- Speech recognition
  - sound input x_t ← observed, x_t is the sound heard at time t
  - Words output z_t ← desired, z_t is the word I said at time t
- Object tracking
  - video input x_t
  - object position output z_t

$$x_{a:b} = x_a, x_{a+1}, \ldots, x_{b-1}, x_b$$

$\arg\max_{z_k} P(z_k \mid x_{1:t})$ where $k > t$

$\arg\max_{z_t} P(z_t \mid x_t)$

$\arg\max_{z_t} P(z_t \mid x_{1:t})$

$\arg\max_{x_1} P(x_1)$

$\arg\max_{z_t} P(z_t \mid x_{1:T})$

$$P(X_1, X_2, \ldots, X_T, Z_1, Z_2, \ldots, Z_T)$$

X_t is observation at time t

Z_t is desired "output" at time t

First-order Markov

$P(Z_t \mid Z_{t-1})$ ← 10^{10} T

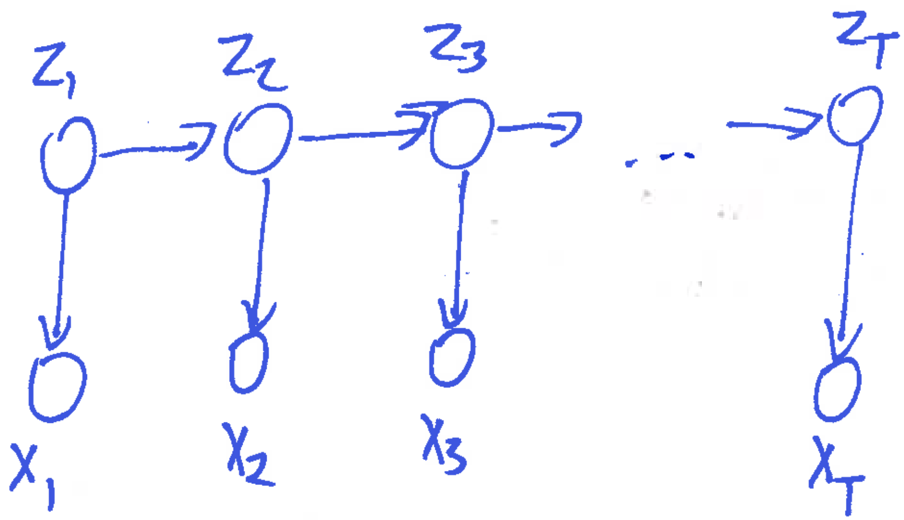
$Z_t \in \{ 10^5 \text{ values} \cup \text{O.O.V.} \}$ unk

$P(Z_T \mid Z_1, Z_2, \ldots, Z_{T-1})$ ← 10^{5T}

$P(Z_T \mid Z_{T-1})$ it's time to eat lunch at ____

Second-order Markov $P(Z_t \mid Z_{t-1}, Z_{t-2})$ 10^{15}

Stationary $P(Z_t \mid Z_{t-1})$ same $\forall t$



Hidden Markov Model

thing

next

$P(X_1, X_2, \dots, X_T, Z_1, Z_2, \dots, Z_T)$

$P(Z_t \mid Z_{t-1})$, assumed same $\forall t$

$P(X_t \mid Z_t)$, assumed same $\forall t$

↳ distribution over sound given word being said

"sensor Markov assumption"

X_t only depends on Z_t

$P(X_t \mid Z_{t-1}, Z_t)$

next thing